DOCUMENT RESUME

ED 094 128 CE 001 623

AUTHOR Crowder, N. David
TITLE A Critique of Duncan's Stratification Research.
SPONS AGENCY Manpower Administration (DOL), Washington, D.C. Office of Research and Development.
REPORT NO DLMA-91-35-68-15-1
PUB DATE 71
NOTE 46p.; Ph.D. Dissertation, Duke University

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; Critical Path Method; Data Analysis; Doctoral Theses; Economic Research; Educational Economics; *Income; Labor Force; Mathematical Concepts; *Mathematical Models; Predictor Variables; *Socioeconomic Status; *Statistical Analysis; Surveys
IDENTIFIERS Duncan (Otis Dudley)

ABSTRACT

Duncan's interpretation of his research on the nature and process of the distribution of rewards in the U.S. stratification system is questioned. Statistical and conceptual flaws are noted in Duncan's work and their implications explored. A reanalysis of primary survey data is presented and, based on a reinterpretation of this data, a reconceptualization of the reward distribution process is offered. The implications of this reconceptualization for stratification research are noted. It is further suggested that a plausible source of Duncan's research error is that of presuppositional bias, particularly American egalitarian achievement ideology. It is also suggested that this presuppositional bias did not emerge as analytical error in Duncan's research as an isolated case, but is present in every functionalist theoretical statement about the stratification system. (Author)

ED 094128

BEST COPY AVAILABLE

CE

A CRITIQUE OF DUNCAN'S STRATIFICATION RESEARCH

by

N. David Crowder

Department of Sociology and Anthropology
Duke University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Sociology and Anthropology
in the Graduate School
of Duke University

1971

BEST COPY AVAILABLE

CE001623

ACKNOWLEDGEMENTS

This report was prepared for the Manpower Administration, U.S. Department of Labor, under research and development grant No. 91-35-68-15. Since grantees conducting research and development projects under government sponsorship are encouraged to express their own judgment freely, this report does not necessarily represent the official opinion or policy of the Department of Labor. The grantee is solely responsible for the contents of this report.

The basic data source and most of the data preparation for this study was provided by the U.S. Bureau of the Census. Without the cooperation and assistance of Benjamin J. Tepping, Director of the Center for Research in Measurement Methods at the Bureau of the Census, the data set could not have been obtained.

Alan Kerckhoff has been helpfully critical of my ideas and data interpretations. Joel Smith has been available, helpful and very tolerant for a long time.

Copyright by

N. David Crowder

1971

Reproduction by the U.S. Government in whole
or in part is permitted for any purpose

BIBLIOGRAPHIC DATA SHEET	1. Report No. DLMA 91-35-68-15-1	2.	3. Recipient's Accession No.
4. Title and Subtitle: A Critique of Duncan's Stratification Research			5. Report Date: January 28, 1974 6.
7. Author(s): N. David Crowder (Dep't. of Sociology, West Virginia University, Morgantown, W.Va. 26505)			8. Performing Organization Rept. No.
9. Performing Organization Name and Address: Duke University Durham, North Carolina 27706 (Dep't. of Sociology)			10. Project/Task/Work Unit No. 11. Contract/Grant No. DL-91-35-68-15
12. Sponsoring Organization Name and Address: U.S. Department of Labor Manpower Administration Office of Research and Development 1111 20th St., N.W., Washington, D.C. 20210			13. Type of Report & Period Covered: FINAL 14.

15. Supplementary Notes: A condensation of this report has been published as "A Critique of Duncan's Stratification Research" in the Jan.,1974 issue of ~~Sociology, the journal of the British Sociological Association.~~

16. Abstract

Duncan's interpretation of his research on the nature and process of the distribution of rewards in the U.S. stratification system is questioned. Statistical and conceptual flaws are noted in Duncan's work and their implications explored. A reanalysis of primary survey data is presented and, based upon a reinterpretation of this data, a reconceptualization of the reward distribution process is offered. The implications of this reconceptualization for stratification research are noted. It is further suggested that a plausible source of Duncan's research error is that of presuppositional bias, particularly American egalitarian achievement ideology. It is also suggested that this presuppositional bias did not emerge as analytical error in Duncan's research as an isolated case, but is present in every functionalist theoretical statement about the stratification system.

17. Key Words and Document Analysis. 17a. Descriptors

Earnings
Education
Mathematical models
Socioeconomic Status
Statistical Analysis
Surveys

17b. Identifiers/Open-Ended Terms

Inequality
Labor Force
Path Analysis
Regression Analysis
Stratification

17c. COSATI Field/Group 5K

18. Availability Statement: Distribution is unlimited. Available from National Technical Information Service, Springfield, Va. 22151.	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 101
	20. Security Class (This Page UNCLASSIFIED	22. Price $3.00

ABSTRACT

Contemporary stratification research, particularly that of Otis Dudley Duncan, is found to contain both conceptual and statistical error. This paper comprises a critique of that research and proceeds on two levels.

First, the statistical base of Duncan's linear models of the socioeconomic life cycle relating educational attainment, socioeconomic status and income of members of the U.S. civilian labor force is reexamined. Path regression analysis is the statistical technique utilized for estimating the relative impact of education and socioeconomic status on income. In this study, these computations are made for a national sample of the U.S. population. It is found that the model is capable of explaining at most 25% of the variance in income and is, therefore, an inadequate characterization of the reward distribution system of the U.S. occupational structure, contrary to Duncan's claims. Contingency tables relating the model variables are examined and it is noted that not only is there wide dispersion of income of persons in each educational attainment category (although income is approximately normally distributed in each category), but there is also the anomaly of sizeable groups of persons with high levels of education and low levels of income; in particular, persons with a high school education and annual income below $4,000 are examined. Their characteristics are elaborated in some detail, but no explanation consistent with the Duncan linear model is found. It is suggested that these low earners represent an extreme

contradiction to the hypothesis of the linear model and that they highlight the inadequacies of that model. The assumption of linear relationships among model variables is examined through an investigation of regression residual plots. The relationships are found to be non-linear although the linear model continues to estimate the non-linear model parameters quite well. The general conceptual implication of this data analysis is that the linear model is inadequate. It is further suggested that one of the conceptual reasons for its inadequacy is that it fails to consider persons and institutions which exercise more or less direct control over the distribution of income rewards. It is also suggested that future research ought to at least acknowledge the existence of this "system" of persons and institutions even if it does not directly consider them for analysis. In the light of this reconceptualization, Duncan's models are seen to be sub-models which articulate the relative impact of educational attainment and socioeconomic status on a part of income achievement.

Second, ideological bias in functionalist theories and in Duncan's research interpretations is noted and it is proposed that this bias led to the faulty conceptualization of the income distribution system and the rationalization of large amounts of unexplained variance noted above.

The work suggests some specific directions for future stratification research. Job performance, as measured by those who control income distribution, as well as normative conformity or the lack of it by income recipients may be examined as potential explainers of income variation. This implies the need for surveys of income recipient attitudes as well as the acquisition of information, demographic and attitudinal, about persons in direct control of income distribution. Low earner and high earner anomalies may be examined in greater detail. Finally, the synchronization

of educational institutions and occupational institutions should be examined; for example, the ratio of the number of persons with a particular educational achievement level to the number of jobs available which pay incomes implied by the linear model for that educational attainment level; and the number of occupations at given income levels. This examination would permit heightened understanding of the relationship between structural or systemic conditions and individual achievement.

Finally, the reconceptualization and the research suggestions render it possible to combine the empirical research of the Marxists (notably on the unevenness of overall income distribution) and the functionalists (including Duncan's research); i.e., to begin to view these disparate research investigations as complementary rather than contradictory efforts.

A CRITIQUE OF DUNCAN'S STRATIFICATION RESEARCH

THE THEORETICAL CONTEXT OF DUNCAN'S RESEARCH

Otis Dudley Duncan's research on the U.S. stratification system spans nearly two decades. The major substantive focus of the work has been upon the relationships among family background, educational attainment and occupational prestige and their relative impact upon the determination of the nature of the U.S. stratification system. The most important methodological contribution of the work is the sophisticated utilization of correlation, regression and path analysis techniques to construct models depicting the operation of the stratification system. The research has been, to a large extent, empirically oriented with little theoretical exegesis beyond the statements offered about the conceptual paradigm of the constructed models. However, in the past few years, Duncan has begun to develop theoretical statements about the stratification process.[1]

Duncan's work has become a major force in present-day American sociological thought on the process of stratification and its theoretical context is that of the functionalist theories of stratification. I find a persistent ideological bias in these functionalist theories beginning at least with the publication of "Some Principles of Stratification" by Davis

and Moore;[2] moreover, this presuppositional bias appears to have caused a general conceptual error in all such theories. That this bias and the conceptual error emerge explicitly in Duncan's work is the thesis I shall develop in the remainder of this paper. It is important to note that the error I find in Duncan's research has not occurred as an isolated case; it is a continuation of a theme that is at least twenty five years old.

Nearly twenty years ago, Tumin responded critically to the functionalist theory of stratification presented by Davis and Moore. In that critique, Tumin suggested that part of the Davis and Moore interpretation of the stratification system is presuppositional, a result of ideological bias.[3] And he was led to reject the Davis and Moore claim that:

> Social inequality is thus an unconsciously evolved device by which societies insure that the most important positions are conscientiously filled by the most qualified persons.[4]

Implicit in the Davis and Moore statement above is the assumption of a benign system, or a system mutually agreeable to all members of a society, which dispenses rewards according to some systemic need. This assumption is repeated, in some form, in every functionalist statement since the Davis and Moore article; sometimes it is labelled the "social system," sometimes it is called the "normative order," and still other times it remains implicit. It represents the constant refusal to depict the system as anything but a static hierarchical structure of postitions with assignation of rewards determined by "society's needs." I wish to reassert that the social system is peopled and reassert that there are multiple normative orders; moreover,

some persons are in a position to impose particular normative orders on others; that is, systemic power is unevenly distributed among persons and their respective institutions.

This section has been omitted.

The following selections offer the clearest illustrations of the impact of ideology upon Duncan's theoretical and policy statements regarding the U.S. stratification system:

> In a liberal democratic society we think of the more basic principle as being that of achievement. Some ascriptive features of the system may be regarded as vestiges of an earlier epoch, to be extirpated as rapidly as possible. Public policy may emphasize measures designed to enhance or to equalize opportunity - hopefully, to over-

come ascriptive obstacles to the full exercise of the achievement principle.

The question of how far a society may realistically aspire to go in this direction is hotly debated, not only in the ideological arena but in the academic forum as well. Our contribution, if any, to the debate will consist largely in submitting measurements and estimates of the strength of ascriptive forces and of the scope of opportunities in a large contemporary society. The problem of the relative importance of the two principles in a given system is ultimately a quantitative one. We have pushed our ingenuity to its limit in seeking to contrive relevant quantifications.[6]

And, similarly:

The basic assumption underlying these conjectures is that a fundamental trend toward expanding universalism characterizes industrial society. Objective criteria of evaluation that are universally accepted increasingly pervade all spheres of life and displace particularistic standards of diverse ingroups, intuitive judgements, and humanistic values not susceptible to empirical verification. The growing emphasis on rationality and efficiency inherent in this spread of universalism finds expression in rapid technological progress and increasing division of labor and differentiation generally, as standards of efficiency are applied to the performance of tasks and the allocation of manpower for them. The strong interdependence among men and groups engendered by the extensive division of labor becomes the source of their organic solidarity, to use Durkheim's term, inasmuch as social differentiation weakens the particularistic ingroup values that unite men in common bonds of mechanical solidarity. The attenuation of particularistic ties of ingroup solidarity, in turn, frees men to apply universalistic considerations of efficiency and achievement to ever-widening areas of their lives.

Heightened universalism has profound implications for the stratification system. The achieved status of a man, what he has accomplished in terms of some objective criteria, becomes more important than his ascribed status, who he is in the sense of what family he comes from. This does not mean that family background no longer influences careers. What it does imply is that superior status cannot any more be directly inherited but must be legitimated by actual achievements that are socially acknowledged. Education assumes increasing significance for social status in general and for the transmission of social standing from fathers to sons in particular. Superior family origins increase a son's chances of attaining superior occupational status in the United States in large part because they help him to obtain a better education, whereas in less industrialized societies the influence of family origin on status does not seem to be primarily mediated by education. Universalism also discourages discrimination against ethnic minorities, though it

> does not furnish incentives for giving them the assistance they may need to overcome the handicaps produced by long periods of deprivation and suppression. At the same time, universalism fosters a concern with materialistic values at the expense of spiritual ones; an interest in achievement and efficiency rather than religious devotion, philosophical contemplation, or artistic creation; a preoccupation with the outward signs of success and little patience for probing the deeper meanings of life. The crass materialism and invidious striving for status in today's world that have often been deplored are an integral part of the universalistic system that has also helped produce many things we cherish, including technological progress, a high standard of living, and greater equality of opportunity.[7]

And, finally:

> Some readers will undoubtedly be disappointed that the new models exhibited here do not result in major increments to the amount of variance "explained" by comparison with the basic model of occupational achievement with which we began. It is implicit in the discussion by Lipset and Bendix (1959, Chapter IX) that a systematic consideration of "Intelligence and Motivation," in juxtaposition with the several sociological variables implicated in the process of stratification, should result in a more nearly complete "explanation" of occupational mobility. Our substantial effort to treat such variables systematically implies agreement with the spirit of their discussion. We were not, however, concerned to move the coefficient of determination much closer toward the asymptote of unity. Instead, we expected to achieve a more thorough understanding of relationships that were already well established, and thus to secure an improved "explanation" in a sense rather different from that conveyed by the magnitude of the multiple correlation. The final judgement of our success is, of course, to be made by the reader; but we would ask that he take as his criterion the cogency of the models and the arguments supporting them rather than the purely statistical norm. There are reasons for believing -- indeed, for hoping, in our capacities a members of a relatively open society -- that nothing like "complete" explanation of occupational achievement will be secured with variables of the kind we now know how to measure. But there is still a long way to go in providing a consistent and convincing structure for the knowledge we already have.[8]

As I will show, one could not infer or build this ideology from the contingency tables relating income, educational attainment and socioeconomic status. It must have preceded the data analysis. The ideology, of course,

is not limited to Duncan. In the preface to Indicators of Social Change, a recent text by Sheldon and Moore containing an article on stratification by Duncan, the authors write:

> The intellectual system, once and perhaps still currently the guardian of tradition and values, has provided one of the integrative functions of our society. In the future as the primary source of innovation and thereby bearer of change, it begins (haltingly) to replace the economy in carrying out the adaptive functions of the society. Education as the purveyor and distributor of knowledge is becoming the major determinant of the stratification system. For better or for worse, our society is beginning to place almost exclusive reliance on educational attainment as the sorting mechanism for adult occupational position.[9]

An error, originally generated by presuppositional bias, is here transformed and presented as fact. The concluding claims above for education are demonstrably false.

DATA, MODEL AND MATHEMATICS

Data Documentation. The primary source of data for the subsequent analysis is the Content Evaluation Study of the 1960 Census conducted by the U.S. Bureau of the Census following the actual census enumeration and considered more accurate than the actual enumeration.[10] There is little need to document the quality of the data in terms of collection, coding, tabulation, editing and compilation. Suffice it to say that it is of the same quality as other Census surveys, notably the U.S. Current Population Survey, and thus is quite comparable in overall quality to the Occupational Changes in a Generation (OCG) data which is the basis of most of Duncan's empirical work. The CES sample used here is much smaller than Duncan's OCG sample but none of the model parameters presented here are significantly altered by this difference; moreover, none of the inferences made here are in any way a product of the difference between the two sample sizes.[11]

The standard population used throughout this study consists of urban, white, male, married, native-born heads of households, between the ages of 20 and 64, who were in the U.S. experienced civilian labor force in 1959. All other populations considered are subsets of this standard population and are noted where used. Otherwise, inferences refer to this standard population.

The Model. The conceptual paradigm for Duncan's model of the socio-economic life cycle process is simply that a man begins his life and spends the early part of it in a family, acquires education, translates that education into an occupation which, in turn, provides income and status

rewards. These life cycle events occur in more or less the temporal order in which they are mentioned above and factors at each prior stage influence variables at the next stage. The sequence of events and the relationships among model variables are given in Figure 1. Duncan has described this model,

[FIGURE 1 ABOUT HERE]

in part, as follows:

> . . . With respect to this model (though not necessarily with respect to other models one might entertain), two measures on a respondent's family of orientation are taken to be "predetermined variables"--that is, the model says nothing about how values of these variables are themselves determined. These two measures of "family background" (which term may serve as a convenient label) are the educational attainment and the occupational status of the head (normally the respondent's father) of the family of orientation. It is supposed that the size of the family of orientation, measured by the number of siblings of the respondent, depends on the two predetermined variables as well as on other factors that are not specified in the model and that are taken to be uncorrelated with the predetermined variables. Further, it is suggested that the respondent's educational attainment depends on how many siblings he has as well as on the two measures of family background and unspecified residual factors. The achieved occupational status of the respondent, as of the time information on him is collected, is taken to depend on prior educational attainment, on number of siblings, on family background, and on unspecified residual factors. Finally, the current money income of the respondent is represented as a function of his occupational status, his educational attainment, the number of siblings he has, and the two measures of family background, as well as unspecified residual factors. The "residual factors," in each case, are the closest approximation we have to an operational counterpart to Schorr's "pure luck." To the extent that future research renders some of the presently "unspecified residual factors" specified, the apparent role of pure luck will diminish with the incorporation of additional specific factors into new models. How much the residual can be made to shrink in this fashion can only be conjectured; all experience with comparable problems suggests that models in the foreseeable future will continue to require substantially weighted terms for pure luck.[12]

I shall be concerned here with the last three variables in this model, as given diagrammatically in Figure 2. Two recursive linear equations

[FIGURE 2 ABOUT HERE]

suffice to mathematically represent this abbreviated model:

(1) $Y = p_{YU}\,(U) + p_{Y\mathrm{Res}Y}\,(\mathrm{Res}Y)$

$H = p_{HY}\,(Y) + p_{HU}\,(U) + p_{H\mathrm{Res}H}\,(\mathrm{Res}H)$, where U stands for respondent's education, Y for respondent's SES, H for respondent's income, and ResY and ResH for the appropriate residual factors.

In terms of the explanation of variation in income, it should be noted that little is lost in the abbreviated model for only one of the excluded variables, head's (father's) occupational status (X), is known to have any direct impact upon income variation, but that impact is relatively small in terms of overall incremental increase in income variation explained.

Path Regression Analysis. Path analysis (the process of determining the parameters of the equations in (1) above) is, mathematically, little more than multiple regression analysis performed on variables subjected to a standard linear transformation. This transformation is a common one attained by first subtracting from the original variable its mean and dividing the resulting quantity by the standard deviation of the original variable; that is, if X is the original variable, then $X^* = \frac{X - \bar{X}}{\sigma_X}$ is the standard variable to which the path coefficients refer. Any variable subjected to the above transformation has mean, $\mu = 0$, and standard deviation, $\sigma = 1$.

Thus, path coefficients are identical to the standardized regression coefficients in a least squares model. In particular, for the abbreviated model of the socioeconomic life cycle given in (1) above, it can be shown that, with the exception of the residual path coefficients, the path coefficients are identical with the standardized regression coefficients obtained from two regressions: i) the regression of socioeconomic status (SES), Y, on educational attainment (EDUC), U, and ii) the regression of income (INC), H, on socioeconomic status and education.[13]

Although the path model is a conceptual tool and not equivalent to the least squares regression model as such, any statistical statements about the parameters (paths) depend upon the regression model; for example, "p_1 is twice as large as p_2" is a statement about the relative power of the two path coefficients to account for variance in the dependent variable.[14]

There are three assumptions that usually underlie inferences made from regression analysis, two of which are essential to it: i) that the error in the dependent variable is distributed with zero mean and variance, σ (unknown), at every value of the independent variables; i.e., the variance of error is the same at each value of the independent variables--this is the homoscedasticity assumption; ii) that errors are uncorrelated; iii) that errors are normally distributed in addition to (i) and (ii). This last assumption is only necessary when establishing confidence intervals for the regression coefficients and performing other statistical tests involving the normal distribution. If the first two assumptions are met, the third one tends to be met also.[15]

DATA ANALYSIS

Variance Explained. In regression analysis, per se, there are two separate but not unrelated problems. One is the truth or the fit of the model; i.e., whether or not linear regression is the appropriate model. The second problem, given the appropriateness of the model, is the relative adequacy of the model as determined by the amount of variation in the dependent variable explained (accounted for) by the model.[16] I shall subsequently argue that, in fact, the regression of income on education and/or socioeconomic status is not linear and thus, that the linear model is technically inappropriate; but their is useful information to be found in an initial assumption of appropriateness of model and an examination of its parameters and variance explained. Now Duncan has argued that:

> Sociologists are often disappointed in the size of the residual, assuming that this is a measure of their success in "explaining" the phenomenon under study. . . .Thinking of the residual as an index of the adequacy of an explanation gives rise to a serious misconception. It is thought that a high multiple correlation is presumptive evidence that an explanation is correct or nearly so, whereas a low percentage of determination means that a causal interpretation is almost certainly wrong. The fact is that the size of the residual (or, if one prefers, the proportion of variation "explained") is no guide whatever to the validity of a causal interpretation. The best-known cases of "spurious correlation" - a correlation leading to an egregiously wrong interpretation - are those in which the coefficient of determination is quite high.[17]

There are logical, statistical and semantic errors in the above statement. Woven together, they produce a fallacious argument and an unwarranted conclusion.

With regard to the (presumed linear) relationships among variables

in the model, there are only two possibilities concerning spuriousness for each relationship: the relationship is either spurious or it is not. Now, if any of the variable associations in the model were thought to be spurious, we would investigate the possibility and immediately discard any segments of the model found to contain spurious associations. But have we not already decided that the relationships between education and income, education and socioeconomic status, and between socioeconomic status and income are not, in fact, spurious? Of course we have.[18] Moreover, all parameters of the model as well as the nature of the model pertain to and depend upon the analysis of variance. All statistics of the model directly relate to variances or ratios of variances. The standardized regression coefficients (or path coefficients) are no more than measures of the extent of variation in the dependent variable given incremental variation in the independent variables and their (the path coefficients) relative magnitude reflects the extent to which each is better (or worse) than the other at the prediction of variation in the dependent variable. If socioeconomic status did not explain more income variance, independent of education, than education explained, independent of socioeconomic status, then the path coefficient from socioeconomic status to income would not be larger than the path coefficient from education to income.[19] Similarly, the coefficient of determination is a measure of the amount of variation in the dependent variable attributable to the separate and combined influences of the known independent variables (non-residual paths) and is thereby an estimate of their importance in the model vis-a-vis all unknown independent variables (residual paths); thus, the coefficient of determination <u>is</u> a measure of the adequacy of the model.[20]

It is impossible to utilize regression analysis and ignore, statistically or conceptually, the amount of variance explained in the dependent variable. We are, in fact, always analyzing variance in the dependent variable with such models. Regression (or correlation) analysis is not just a convenient computational device for path analysis--it is absolutely essential to it so long as the linear equations of path analysis are not perfectly fitted to the data; that is, so long as there is variation in the dependent variable about the fitted line. Insofar as this variation exists, we are fitting lines to data by the process of least squares and when we make statements about the statistical parameters of that process, we are subject to the rules of regression analysis.[21]

Assuming that the reader is now convinced of the merits (indeed, the necessity) of the consideration of variance explained in any regression model, let me now turn to the model parameters for the standard population of the CES sample and a consideration of the variation in total income explained by that model. These parameters are given in Table 1. The multiple

[TABLE 1 ABOUT HERE]

correlation coefficient for the complete model is $R = .405$ which implies that total variation in income accounted for by the model is 16.4% ($R^2 = .1640$). The obvious corollary is that 83.6% of the variation in income is left unexplained by a linear regression of income on socioeconomic status and educational attainment. Thus, while educational attainment and socioeconomic status may be the best pair of predictors of variation in income, they nevertheless predict poorly.

Contingency Tables. The observation that the model accounts for only 16% of the variance of the dependent variable, income, raises some obvious questions. What are the patterns of distribution of educational attainment with respect to income, socioeconomic status with respect to income, and educational attainment with respect to socioeconomic status that account for this low level of predictive power in the linear model? Are there any special subpopulations whose income falls so far off the regression line (outliers) as to distort the overall predictive validity of the model? If so, what are the changes in predictive power of the model when they are removed? Let me consider the questions in the order presented beginning with the distribution patterns of each model variable over the standard population. Bar graphs of these distributions are presented in Figure 3 where the height of the bar represents approximate percentage of persons in the interval enclosed by the width of the bar. Note that income, in the trun-

[FIGURE 3 ABOUT HERE]

cated range $0 to $12,000, which includes approximately 94% of the total standard population, is approximately normally distributed. Approximately 51% of the standard population is found in the three largest categories of educational attainment--8, 12 and 16 years of school completed--with twelve years of school completed the modal educational attainment group, containing 28.4% of the total standard population. The modal socioeconomic status group is the 10-20 interval containing 25% of the population; the remaining ten point SES intervals range from just over 7% to just over 12% up to SES = 80.

The highest two SES intervals fall somewhat below this 7% - 12% range. Thus, socioeconomic status is moderately evenly[22] distributed for values of SES below 80, except for the modal SES interval.

Let me turn now to a consideration of the cross classification of each pair of model variables. The correlation (regression) coefficient for a pair of variables provides a summary statement of their relationship and, thus, a summary of the configuration to be found in the cross classification of the two variables. But many particular distribution patterns could produce the same correlation (regression) coefficient; therefore, the contingency table is not merely an alternative to the coefficient of association, it is an adjunct to it. Each provides unique and complementary information about the relationship under investigation. The relevant contingency tables are contained in Tables 2, 3 and 4.

[TABLES 2, 3 AND 4 ABOUT HERE]

The distribution of income over each ten point SES interval is approximately normal with median incomes of SES categories ranging from approximately $4,000 in the lowest ten point SES interval to approximately $8,500 in the highest interval. Implicit in the nearly normal distribution of income is that there is some concentration of persons about the mean income; however, the range of the distribution in every ten point SES interval is at least $7,000 wide. This general pattern of the distribution of income over socioeconomic status is maintained even within smaller ranges of the educational distribution.

There is a fairly wide range of distribution of socioeconomic status

in each of the three largest educational attainment categories. In particular, approximately 94% of the persons in the eight years of school completed category are distributed across the 0-60 SES range with about 33% of the category total contained in the 10-20 SES interval and the remainder spread somewhat evenly over the 0-60 SES range. For the twelve years of school completed category, the pattern is similar but covers a wider range of SES, 0-80, with approximately 22% of the educational category total contained in the 10-20 SES interval. The college graduate population (sixteen or more years school completed) is distributed over the high socioeconomic range (SES greater than 50). The distribution of socioeconomic status in other educational attainment categories tends to approximate these patterns.

The cross tabulation of educational attainment and income is the most illuminating contingency table under consideration. The most salient single observable pattern is that of the approximately normal distribution of income in each educational category with moderately rising mean incomes with increasing educational attainment levels. The distribution is always over a large income range, especially in the case of the modal educational level, twelve years of school completed. In particular, for the eight years of school completed category, the distribution of persons is approximately normal in the $0 - $10,000 income range. For the twelve years of school completed category, the pattern is similar but covers a wider range of income: $0 - $12,000. For the sixteen years of education completed category, the population is distributed over the $4,000 - $13,000 income range but there is less evidence of a normal distribution.

While it is true that there is an increase in average income with increasing years of educational attainment, the incremental increase is

small in comparison to any of the income ranges mentioned above. For example, in the standard population, the median incomes for the three largest educational attainment categories--8, 12 and 16--are approximately $5,000, $5,800 and $7,500 respectively; and incremental increases in median incomes for other educational attainment values are approximately even except for the jump between thirteen and fourteen years which is from $6,000 to $7,000. Thus, the popular notion (to some extent supported by previous interpretations of the model under investigation here), that those persons with a high school education earn more money than those who do not complete high school, cannot be accepted unless qualified by the above facts. A more accurate statement of the reality, it seems to me, is that while a high school education does not guarantee higher income, it does offer the opportunity for some to earn higher wages; i.e., the range of incomes for high school graduates is wider (and the upper bound is higher) than those of the lower educational attainment categories, notably the eight years completed category.[23] There are however large numbers of high school graduates earning no more (in many cases, less) than many persons with less education. Nevertheless, there is a pervasive myth (based, as all myths are, on some truth) both among the general public and some students of the income distribution process that increasing education means increasing income and these wide ranges of income per educational attainment category are either not seen or ignored. The myth provides only a partial answer to the question of the nature of the reward distribution system. To the extent it is not tempered by the facts presented here, the myth represents a serious distortion of the reward distribution process.

The second question raised by the fact of low amount of variance explained by the path model was whether or not there are particular sub-populations which somehow distort the model in question. The general answer to such a question, prior to any research, is of course affirmative; for, in selecting the population for study, we have already eliminated sub-populations which are known or thought to be rewarded by a different process (or rate) than that of the hypothesized model; for example, women and non-whites. An additional group of persons that seemed of potential interest in this regard was that group comprised of persons with high educational attainment, say high school graduates, but with low levels of income; in particular, those persons with twelve years of school completed who were earning less than $4,000 annual income. In the standard population, approximately 16% of the high school graduates are in this income category. Since the model will not predict <u>any</u> income below $4,100 for persons with twelve years of school completed, this sub-population potentially represents a genuine model anomaly.

In the process of examining this highly educated low income group, I created two additional homogenizing constraints on the standard population. I thought perhaps these persons were disproportionately part-time workers, so I reduced the sub-population to those working full time (48-52 weeks). Moreover, it was suggested by contingency tables of age and income that this group was disproportionately young and disproportionately old, so I reduced the full time worker sub-population further to include only those between the ages of 30 and 54.

Table 5 contains the cross classification of education and income for low earners of three standard population subsets: (a) full time workers,

(b) workers between 30 and 54 years of age and (c) full time workers between 30 and 54 years of age. Whereas the total standard population is reduced by

[TABLE 5 ABOUT HERE]

about 41% from the combined effects of age and work status truncation, the total low earner sub-population is reduced by 63% from the same truncation.[24] So that we may say, consistent with a priori expectations, that the overall incidence of low income is, in part, attributable to age and work status. However, the twelve years of school completed subset of low earners is reduced at approximately the same rate as the total low income sub-population.[25]

Thus, while age and work status provides a partial explanation of the existence of some of the low income population, the anomaly of low income high school graduates is not eliminated. These persons remain an extreme contradiction, though a small proportion of the total study population, to the hypothesis of the linear model.[26]

One important by-product of my investigation of the low income anomaly was the computation of all model parameters for the age and work status truncated sub-populations. Noting only the extreme example, for the subset population including only full time workers between the ages of 30 and 54, explained variance in income is increased from 16.4% to 18.5%. Thus, age and work status combined make only a slight difference in the overall adequacy of the model to account for income variation. However, the relative magnitudes of the paths are noticeably altered in this truncated sub-population. The path from education to income is raised from .13 in the standard population to .21 in the age and work status truncated subset while the path from socio-economic status to income is lowered from .32 to .28 in the same populations.

Thus it seems clear that, in addition to Duncan's cohort analysis solution to the problem created by age differentials, any regression based models utilizing an income variable ought to include only full time workers.[27]

The major empirical point to be emphasized in this segment of the analysis, which began with the search for anomalies in the study population and the desire to assess their impact on the adequacy of the model, is that extreme homogenization of the study population leads to a very small incremental increase in income variance explained and the contingency table patterns noted earlier are, in general, maintained in these truncated populations.

Regression Residuals. To this point, data analysis has been technically based upon the assumption of linearity of the variable relationships in the model. If the model fits; i.e., the relationships are linear (regardless of magnitude); then, statistically, the only analytic concern is with the relative magnitudes of the model parameters and the overall predictive power of the model. I have conducted a visual analysis of the regression residuals[28] obtained from the application of the model to the sub-population of full time workers between 30 and 54 years of age. The major indications of this analysis are that error variance is heteroscedastic (in violation of a necessary regression assumption) suggesting the need for a weighted least squares or curvilinear model; and that possibly the underlying relationships between income and education and between income and socioeconomic status are not linear.[29]

In an effort to obtain some indication of the impact of a curvilinear (or weighted least squares) model upon model parameters and amount of variance accounted for, I applied the model to two different transformations of the dependent variable - $\log_e$ of income and the square root of income. Two im-

portant analytic points emerge. First, each of these transformations produces an increase in the proportion of variance explained, from about 18.5% to about 25% (for full time workers between 30 and 54 years of age), suggesting that a curvilinear (or weighted) model is more nearly appropriate. Second, the linear model estimates moderately well the parameters of the more nearly appropriate curvilinear (or weighted least squares) model;[30] this is the usual basis for justifying linear approximations to curvilinear relationships in addition to the increased computational difficulties of curvilinear models.

CONCLUSION

Conceptual Implications. Throughout the foregoing data analysis there is the implication that Duncan's model is either inappropriate or inadequate. An examination, over several study populations, of the ranges of variance explainable by the model shows that unexplained variance in income is quite high (range: 75% to 84%). Having reestablished the importance of unexplained variance in regression based models, it therefore seems to me impossible to continue to hold to the idea that the model is an adequate characterization of the reward (income) distribution system of the U.S. occupational structure.

Duncan's conceptual strategy has been to claim that the amount of variance explained is not an important determinant of the adequacy of his model. I have insisted, by way of statistical argument, that he has erred in this claim. Moreover, Duncan's error in this respect appears either to have been caused by or to have caused his faulty conceptualization of the U.S. income distribution process. This inadequacy in conceptualization was clearly evident in the patterns of variable relationships observed in the contingency tables and is to some extent discernible in an examination of the regression residuals produced by the model; namely, the model fails to provide an explanation of the fact of wide distribution of income in given educational attainment categories and it fails to acknowledge the conceptual importance of the likely curvilinear nature of variable relationships.

A reinterpretation of the data and a reconceptualization of the income distribution process clearly seems called for. To this end, I would suggest

that the model's large residual paths, which Duncan has on occasion attributed to 'pure luck',[31] are to a large extent produced by a combination of constraints generated by an occupational system (including, most importantly, the distribution of income) which exists independent of and prior to any person entering it whose achieved and familial ascribed characteristics we might measure. Moreover, there are persons, whose characteristics are not measured or included in the model, who act more or less in congruence with these systemic constraints when they hire, fire, promote and demote occupational role incumbents. There are systemic components of control over income distribution known to be independent of any particular income recipient or group of recipients,such as the ratio of particular income level jobs to the number of individuals who ought to obtain that income if the linear model applies. In addition, some of the so-called control variables already in use; e.g., skin color and sex, clearly represent a priori systemic constraints.[32]

In addition to independent systemic constraints, there is of course the possibility that some segment of the unexplained variance is attributable to income recipient characteristics as yet unknown and/or unmeasured.[33] Two candidates for inclusion come to mind: job performance as determined by those who control and distribute income and income recipient's level of conformity to the norms of persons and institutions controlling the distribution of his income.[34] It has been argued recently, in the context of Duncan's more elaborate background models, that conformity to prevailing systemic norms is a factor in the determination of success in educational institutions.[35] Given that educational institutions are, in part, places of preparation for membership in occupational institutions, it seems not unlikely that conformity

in the latter would have some impact upon income achievement. In other words, if conformists in high school disproportionately receive better grades, then it seems at least plausible that conformists on the job will get more money.

To the extent that there are more men than jobs at a given income level for a given educational level, 'pure luck' is no doubt a factor in job acquisition, but it is by no means the characterization of residual paths in the model. If, for example, there are only 500 jobs available which pay the linear model rates for a high school education and there are 5,000 high school graduates, some 4,500 must 'lose' in terms of our linear model. The data, especially the contingency tables, strongly suggest that something like this is, in fact, happening though the actual extent is unmeasured. While a high school education is becoming universal, the synchronization of that process with the 'needs' of the occupational structure remains considerably awry when measured against the standard of the linear model.[36] This is not an argument against the universality of education; it is simply an assertion that increased income does not necessarily follow from the increasing universality of high educational attainment.

The likely non-linearity of the variable relationships in the model strongly suggests that, even within the small segment of income variation attributable to extant model variables, equal increments of increase in independent variables implies not a fixed increment of increase in the dependent variable, but one that increases with the level of the independent variable. Thus, for example, incremental income payoff for a year of education in the high end of the educational attainment range is greater than for an equal increment in a lower educational attainment range. This is not implied, conceptually or statistically, by a linear model unless that

model is presented as an approximation of the more appropriate curvilinear model. Certainly, this problem of curvilinearity becomes crucial when one begins to make inferences from the model about the fairness of the system, equal access to the system, cycles of poverty, and so forth.

In the context of the reconceptualization presented here, what do Duncan's models characterize? It seems to me that they now can be seen as sub-models which direct attention to the issues of ascription and achievement, but which may not be used to characterize the entire reward distribution system since most of the control of the distribution of rewards is completely outside these sub-models and therefore not explainable in terms of achievement and ascriptive characteristics of income recipients. Moreover, even within the sub-models, if by a process of eliminating populations such as young workers, black workers, female workers, old workers, and so forth, we arrive at a population for whom achievement criteria of reward acquisition is more important than ascriptive criteria, it hardly seems reasonable to then assert that the sub-system is achievement oriented. Such a claim, dependent of course on the extent of population homogenization, is not unlike the claim that the U.S. social system is non-racist because it does not discriminate against white anglo-saxon Protestants.

The following provides a summary statement of the recharacterization of the income distribution system implicit in the foregoing reconceptualization and reinterpretation of data: The total amount of income rewards from occupational pursuits is to a large extent beyond the ascriptive or achievement characteristics of any particular person or group of persons in the occupational system;[37] but within this constraint of a more or less fixed total, achievement and familial ascribed characteristics of occupational

role incumbents account for approximately 16% to 25% of income variation while luck (an outcome of systemic constraints), unknown system variables (not attributable to income recipients), and as yet unknown and/or unmeasured income recipient variables account for the remainder.

Some Implications for Future Research. Any research (or theory) which proposes to examine the whole stratification system must necessarily abandon the idea that the nature of that system may be determined by reference to the relative magnitude of ascriptive and achievement factors in the acquisition of income rewards.

Survey research which obtains only demographic information on income recipients is inadequate to operationalize the reconceptualization I have offered. For example, to even begin to investigate the impact of conformity, one needs to have some measure of the variable on income recipients as well as some understanding of the normative system by which conformity is defined. The former requires knowledge of attitudes of income recipients; the latter requires information on persons and institutions which control and distribute income.

Structurally, information is needed about numbers of occupations in particular income ranges and numbers of persons at particular educational levels. The nature and rate of change (if any) of the synchronization of these two institutional facets of the stratification system need to be examined.

Finally, the low earner anomaly provides a good example of the direction some stratification research might take. It seems to me that to focus intensive research effort upon these persons would potentially provide far more understanding of the stratification process than continued demographic

surveys of the entire labor force. Similar attention could of course be directed toward relatively high income populations.

Ideology in Duncan's Research Interpretations. I wish to illustrate the basis of my inference that Duncan's interpretations of his empirical analyses of the U.S. stratification system are ideologically biased; that a part of what is presented as objective interpretation of methodologically sophisticated research is, in fact, presuppositional.

The presence of American egalitarian achievement ideology is not only evident in Duncan's interpretations of his research, it is also apparent in his research focus; e.g., in The American Occupational Structure,[38] there is little or no consideration of income, the primary reward of that structure; moreover, the ideology is evident in his denial of radical, notable Marxist, stratification theorists; and, paradoxically, it is apparent in his discussion of and disdain for the skin color caste system in the U.S.; in particular, in his investigation of the differential applicability of his model to whites and blacks.

I shall try to illustrate these claims through brief annotated quotations from Duncan's work. The selections to follow are not out of context in the sense that if read with the material surrounding them in the source work, they would necessarily mean something else; but they are out of context in the sense that they are quite small extractions from a large volume of work. They nevertheless illustrate my points.

On stratification theorists, Duncan writes:

> Cooley and Sorokin are of permanent value (or as near to being so as any sociological writing can be) for conceptual orientation, and are much more useful for the purpose of acquiring concepts suited to the

study of specifically American social stratification than are the writings of Weber, Veblen, or Marx.[39]

Duncan's major conceptual error, the ignoring of systemic constraint upon reward acquisition, leads directly to labelling a model which represents only a segment of the stratification system as the model of the system. This mislabelling in turn is either caused by or leads to the false conception that familial ascription versus individual achievement is the measurement problem that needs to be solved in order to characterize the system; i.e., the paramount problem of stratification research, and that:

> Evidently the task of social analysis is quite different, according to whether the investigator is concerned with a system with predominantly ascriptive principles or with one in which most statuses are achieved. Assuming that American society falls toward the latter pole, the analyst's task is somewhat like that of one who bets on the races. A stratified society which places stress on achievement is not unlike a race in which the runners differ not only in skill and ability, but also in respect to various advantages or handicaps. Some begin the race with heavy packs upon their backs and many obstacles in their course, while others enjoy freedom from such impediments. The race is rendered less predictable, too, in that these initial conditions may be modified as the contest progresses, not only in response to the success of the contestants up to a given point, but because the handicaps are varied somewhat randomly during the running. The outcome is hardly determinate when the race begins, but an informed bettor could nevertheless make money if his odds were accepted. From a normative standpoint, one can be concerned with whether the race is run fairly, given the rules, or whether the rules themselves should be changed.[40]

An informed bettor, accepting the somewhat limited analogy for a moment, would be one who also had some inside information on which runners the race organizers favored and which ones they would disqualify or add handicaps to and what criteria they use to do so. Moreover, one could bring into question the whole concept of the race and endeavor to discover just why these persons wish to hold these races, who benefits from them and how, and who are the rulemakers and how did they come to acquire this power. The analogy gets

better as we add race organizers (and of course their vested interests), race officials, off and on track bettors, etc. It also gets a bit more Marxist and, thus, a bit closer to the reality which it pretends to illuminate. Duncan's unwillingness to conceptualize in this fashion is exemplified in the statement that: 'If one believed - it is not suggested that anyone does or has cause to do so - that all rewards are a function of ownership of the means of production, he would also have to believe that there is great rigidity in the system of inequality . . . .'[41] The problem, however, is not whether it is all rewards or none--it is to what extent. He gets nearest my reconceptualization when he states:

> The foregoing discussion implies, or presupposes, a concept of "stratification system" Such a system comprises two analytically distinct components, both essential to the concept: (1) a set of one or more hierarchies of institutionalized inequality with respect to the statuses (rewards, evaluations, prerogatives) conferred on the basis of incumbency and performance of roles; (2) a pattern of intergenerational transmission of status or access to roles such that the position of an individual on a scale of inequality is associated to a nonnegligible degree with the position of his family of orientation.[42]

He apparently recognizes that a system exists but he must consider it somehow static, fixed, unpeopled, and uninfluential in the ongoing control and distribution of occupational rewards and/or the assignation of occupational roles. Seemingly, he believes that variation in rewards is either ascertainable from analysis of income recipient characteristics or that it is not ascertainable at all; i.e., he will discuss variation in income when his model variables explain it and will deny its importance when they do not. This not only contradicts the facts but the nature of the statistical techniques used to analyze them. Duncan seems to be incognizant of Marx's point

that the stratification system not only exists independent of occupational role incumbents but also that it is the major determinant of reward distribution.

Now I wish to explicate the seeming paradox of ideological bias in the context of Duncan's explication of and attack upon the nature of black-white differentials in income. In an article entitled 'Inheritance of Poverty or Inheritance of Race?' Duncan concludes that:

> It is true, of course, that in American society one is well advised to "pick his parents" so that he begins life on a favorable socioeconomic level. But the models exhibited here fully support Gallaway's conclusion that this strategy is not nearly so important as previous doctrine would seemingly have us believe. It is, however, of vital importance to choose parents of the "right" skin color if one wants to avoid a high risk of ending up at a low level on the income scale. In general, the supposition that the "poor are poor because they are poor" is not only an intellectual obfuscation, but also a feeble guide to policy in what is obviously the most desperate and refractory sector of the "poverty problem," that is the "race problem."
> I have no doubt that the instigators of the War on Poverty thought that it could be planned in such a way as to remedy the gross discrepancies in achievement and rewards between the races. But this just does not happen as a benign fallout from conventional measures taken to enhance "opportunity." Until we summon up the courage to distinguish between the problems of poverty and the problems of race, we shall have to reckon with the consequences of our lack of candor.[43]

And, in The American Occupational Structure, we find:

> The general conclusion to which these findings point is that the American occupational structure is largely governed by universalistic criteria of performance and achievement, with the notable exception of the influence of race.[44]

There are, in fact, more whites living in poverty than blacks even though the proportion of the black population in poverty is far larger than that of the white population. And, without any doubt, this disproportionate distribution of the incidence of poverty is attributable to skin color

discrimination.

The point I wish to make is that Duncan's cogent argument that blacks in the U.S. are disproportionately poor because of skin color discrimination is not also an argument which refutes the validity of concepts such as 'cycle of poverty' and 'inheritance of poverty.' Suppose, for a moment, that educational and occupational discrimination against black persons were completely eliminated and that blacks were economically assimilated so that the income, education, SES and family background model differences noted by Duncan were no longer discernible in our data. Would the poor disappear? Of course not; the proportion of blacks who were poor would be reduced to the level of incidence of poverty in the white population; but the number of persons living in poverty would remain quite large and unless one is willing to argue that they are persons who have had an equal opportunity to compete in a fair system and have failed (so that they 'deserve' to be poor), the concepts of 'cycle of poverty' and 'inheritance of poverty' remain potentially relevant. Duncan obviously perceives the negative educational, occupational and economic consequences of skin color caste, but he steadfastly refuses to see that the stratification system, per se, is in large measure responsible for the creation and maintenance of poverty, as well as other facets of the income distribution process.

Finally, a point of agreement. Blau and Duncan write, in the preface to The American Occupational Structure, that:

> Confronted by the same set of quantitative data, two men do not necessarily arrive at the same conclusion regarding the empirical "facts" of the case, let alone regarding the inferences to be drawn from them. A configuration clearly apparent in a number of complex tables to one may be seen by the other as conforming to a different pattern, dependent on initial assumptions and problem focus. Orders of significance and priority of emphasis may fail to coincide, and

> what looks like an interesting discovery from one point of view seems trivial from another.[45]

I agree.

NOTES

1. For example, O.D. Duncan, 'Social Stratification and Mobility: Problems in the Measurement of Trend,' in E. B. Sheldon and W. E. Moore (eds.), Indicators of Social Change, Russell Sage Foundation, New York, 1968.

2. K. Davis and W. E. Moore, 'Some Principles of Stratification,' American Sociological Review, Vol. 10, 1945, pp. 242-49.

3. Melvin Tumin, 'Some Principles of Stratification: A Critical Analysis,' American Sociological Review, Vol. 18, 1953, pp. 387-97.

4. Davis and Moore, op. cit., p. 243.

5. Tumin, op. cit., p. 389.

6. Peter M. Blau and Otis Dudley Duncan, The American Occupational Structure, John Wiley and Sons, New York, 1967, pp. 163-64.

7. Ibid., pp. 429-30.

8. Otis Dudley Duncan, David L. Featherman and Beverly Duncan, Socioeconomic Background and Occupational Achievement: Extensions of a Basic Model, U.S. Department of Health, Education and Welfare, Washington, D.C., 1968.

9. E.B. Sheldon and W.E. Moore, Indicators of Social Change, Russell Sage Foundation, New York, 1968, p. 14.

10. U.S. Bureau of the Census, 1960 Census of Population and Housing Eval-

uation and Research Program: Study EP-10: Content Evaluation Study (Population), Washington, D.C., 1962. I am indebted to Dr. Benjamin J. Tepping of the U.S. Bureau of the Census for access to and assistance in the compilation of the CES data used here.

11. A more detailed comparison of the CES and OCG samples is available to the reader upon request.

12. Otis Dudley Duncan, 'Inheritance of Poverty or Inheritance of Race?' in Daniel P. Moynihan (ed.), On Understanding Poverty, Basic Books, Inc., New York, 1969, pp. 89-91.

13. Otis Dudley Duncan, 'Path Analysis: Sociological Examples,' American Journal of Sociology, Vol. 72, 1966, pp. 1-16. Also in this article, Duncan provides the derivation of a method for the computation of residual paths.

14. Non-standardized regression coefficients, on the other hand, provide estimates, in original variable units, of changes in the dependent variable given particular changes in the independent variables. Since the units of measurement differ in kind, no comparison of independent variables is warranted.

15 Norman Draper and Harry Smith, Applied Regression Analysis, John Wiley and Sons, New York, 1966, p. 17.

16. Ibid., pp. 1-32.

17. Blau and Duncan, op.cit., pp. 174-75.

18. All correlations are potentially spurious. We come to believe that they are not on extra-statistical grounds; and we come to believe that they are by means of proof of their spuriousness as well as by way of extra-statistical reasoning. In addition, a spurious correlation does not necessarily lead to 'egregiously wrong interpretation.' Wrong interpretation follows either from ignorance of the implications of spuriousness or from mistaking a spurious correlation for a non-spurious one. Moreover, the coefficient of determination need not be high in the case of spurious correlation any more than it need be high in the case of non-spurious correlation. The 'best-known cases' have high coefficients of determination no doubt because cases of spurious correlation that have high coefficients of determination make better examples than those that do not.

19. This statement is true regardless of the effects upon income of the interaction between educational attainment and socioeconomic status.

20. The residual paths, along with the coefficients of determination, are often relegated to secondary status in Duncan's discussions. But, by the same major criteria he uses to determine exclusion and inclusion of model variables and to determine the relative importance of the included variables, the residual paths are clearly the most important ones in the model. That their constituent variables are unknown does not diminish their statistical or conceptual importance.

21. The fact that our study populations are samples increases the reliance upon the statistical model as a basis for inference.

22. The quantitative descriptive adjectives utilized in describing contingency table patterns are not intended to represent measures of anything. They represent my interpretation of particular patterns which may be more fully understood by direct observation of the tables.

23. Rather than providing more or less direct access to higher rewards, it seems to me that increased education tends to provide access to competition for higher rewards in a system where apparently the proportion of 'winners' (in terms of the difference between the predictions of the linear model and the actual distribution of income) is quite small.

25 ~~24~~. While the combined effects of age and work status truncation are the same (65% versus 63% reduction), their independent effects differ between the two populations. The independent effect of age truncation is noticeably greater among the high school graduate low earners than it is in the total low earner population.

24 ~~25~~. Because of the smallness of some of the cell sizes in Table 5, I constructed similar tables utilizing the 1/1,000 sample of the 1960 U.S. Census--twelve times larger than the CES sample. Judged by the standard of the 1/1,000 sample tables, the percentage distributions contained in Table 5 are accurate. (U.S. Bureau of the Census, Census of Population and Housing: 1960; 1/1,000 - 1/10,000, Two National Samples of the Population of the United States, Washington, D.C., 1960.)

26. Alan Kerckhoff suggested to me that these low earners may be 'ghettoized,' regionally concentrated; e.g., in the southern U.S., earlier married, burdened by large families, etc. In a preliminary effort to

examine these hypotheses about the low earner anomaly, I have used the 1/1,000 census sample (see note 25) to compare low earning high school graduate full time workers between 30 and 54 years of age (called 'low earners' below) to the remainder of the high school graduate full time workers between 30 and 54 (called 'high earners' below) on selected characteristics such as region of residence, size of place of residence, age, size of family, year of first marriage, work status of spouse, etc. Space does not permit a complete report of the details of my analysis, but two points do seem worthy of note. First, the hypothesis of regionalization is to some extent supported: 33% of the 'low earners' live in the southern U.S. whereas only 21% of the 'high earners' live in that region. Second, with respect to residence in Standard Metropolitan Statistical Areas (SMSA), 'low earners' are slightly more concentrated in central cities of SMSA's with populations over 1,000,000: 21.6% compared to 17.3% of the 'high earners;' whereas 15% of the 'low earners', compared to 28% of the 'high earners,' live in the remainder of these large SMSA's. The most interesting SMSA residence differential is that 31% of the 'low earners,' in contrast to 20% of the 'high earners,' live outside SMSA's; moreover, fewer 'low earners' work in SMSA central cities (5.5% / 11.4%) and more 'low earners' work outside the SMSA (25% / 16.7%). Thus, the ghettoization hypothesis seems, in the main, unsupported by the data. The unpublished details of this analysis are available upon request.

27. Work status (weeks worked) and age seem to me to be intrinsically different kinds of variables with respect to the present analysis. If a man with relatively high education and low income is found to be a

part-time worker, we may say that his low income is explained by part-time employment status and it seems reasonable to eliminate such persons from any analysis involving income prediction and still be able to legitimately apply our models to the population at large, so long as the rates of rewarding are not different. In the case of age truncation, it is clear that the rates are different and model applicability would seem thereby restricted.

28. These residuals should not be confused with the residual paths in the model or with the amount of unexplained variance. The regression residual is defined as the difference between the predicted and observed values of the dependent variable. These residuals represent observed error and the plots of residuals against predicted values and against independent variables should provide a pattern of a horizontal band in the case of a linear model. Particular deviations from this standard pattern are indicative of non-linear relationships among model variables and/or the violation of regression assumptions. (Draper and Smith, op. cit., pp. 86-95.)

29. Again, space permits only a summary statement of my research on this topic and again the unpublished details of the analysis, including the residual plots, are available upon request.

30. For example, in the case of the model applied to the square root transformation and full time workers between 30 and 54 years of age, the value of the path from education to the square root of income is .24 and that from socioeconomic status to the square root of income, .33; whereas, in the model containing an untransformed income variable, the respective path values are .21 and .28.

31. Duncan, 'Inheritance of Poverty or Inheritance of Race?' op. cit., pp. 88-91.

32. Surely, the long history of labor-management fights in the United States suggests that there is a system 'out there' which to a large extent controls incomes via persons and institutions which support it.

33. Duncan has shown that both father's socioeconomic status and education have measureable impact upon respondent's socioeconomic status and/or income (Duncan, 'Inheritance of Poverty or Inheritance of Race?,' op. cit., Figure 4-1, p. 90), but that impact is relatively small in terms of overall incremental increase in income variation explained.

34. It is likely that there would be some interaction between these two variables but it also seems likely that each potentially has a discernible impact on reward attainment; in addition, each might be differentially important depending where in a status hierarchy observation of its impact is made.

35. J.N. Porter, 'On Making It: Race, Socialization and Mobility in Educational and Early Occupational Attainment,' unpublished Ph.D. dissertation, Duke University, 1971.

36. An extreme contemporary example is the case of the Canadian province of Quebec which has in the recent past produced sufficiently too many college graduates for the number of college level jobs to make it a nationally newsworthy story.

37. Disputes between labor and management are usually over relatively small

increments and it is always management which relinquishes some of the wealth which it controls and distributes.

38. Blau and Duncan, op. cit.

39. Duncan, 'Social Stratification and Mobility: Problems in the Measurement of Trend,' op. cit., p. 675.

40. Ibid., p. 685.

41. Ibid., p. 689. For a very cogent argument that there is indeed great rigidity in the system of inequality, see Gabriel Kolko, Wealth and Power in America: An Analysis of Social Class and Income Distribution, Praeger, New York, 1969.

42. Duncan, 'Social Stratification and Mobility: Problems in the Measurement of Trend,' op. cit., p. 690.

43. Duncan, 'Inheritance of Poverty or Inheritance of Race?,' op. cit., pp. 108-09.

44. Blau and Duncan, op. cit., p. 32.

45. Ibid., p. ix.

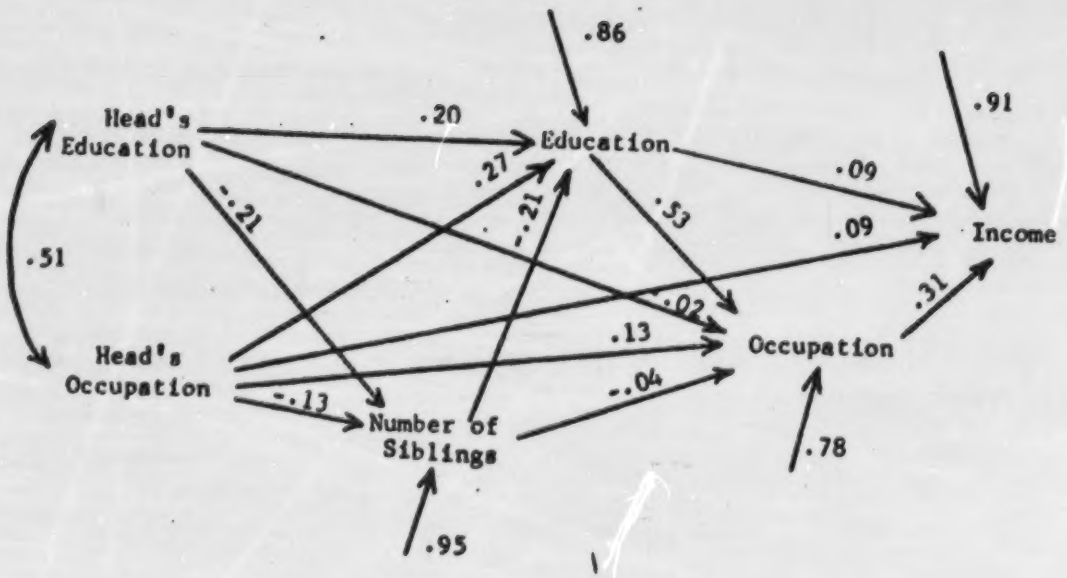

FIGURE 1. Path Diagram of a Model of the Socioeconomic Life Cycle with Path Coefficients Estimated for White Native-Born Men, 25 to 64 Years Old, with Nonfarm Background and in the Experienced Civilian Labor Force: March, 1962. (Source: O.D. Duncan, 'Inheritance of Poverty or Inheritance of Race?,' in Daniel P. Moynihan (ed.), On Understanding Poverty, Basic Books, Inc., New York, 1969, Figure 4-1, p. 90).

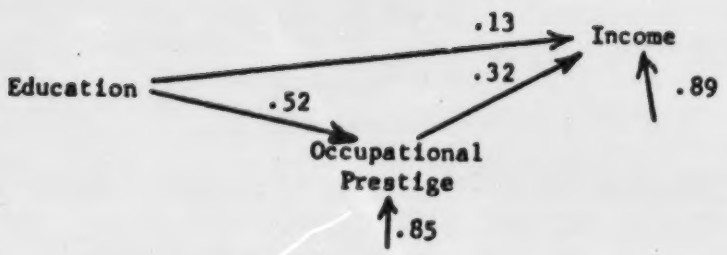

FIGURE 2. Basic Path Model with Path Estimates from CES Standard Population.

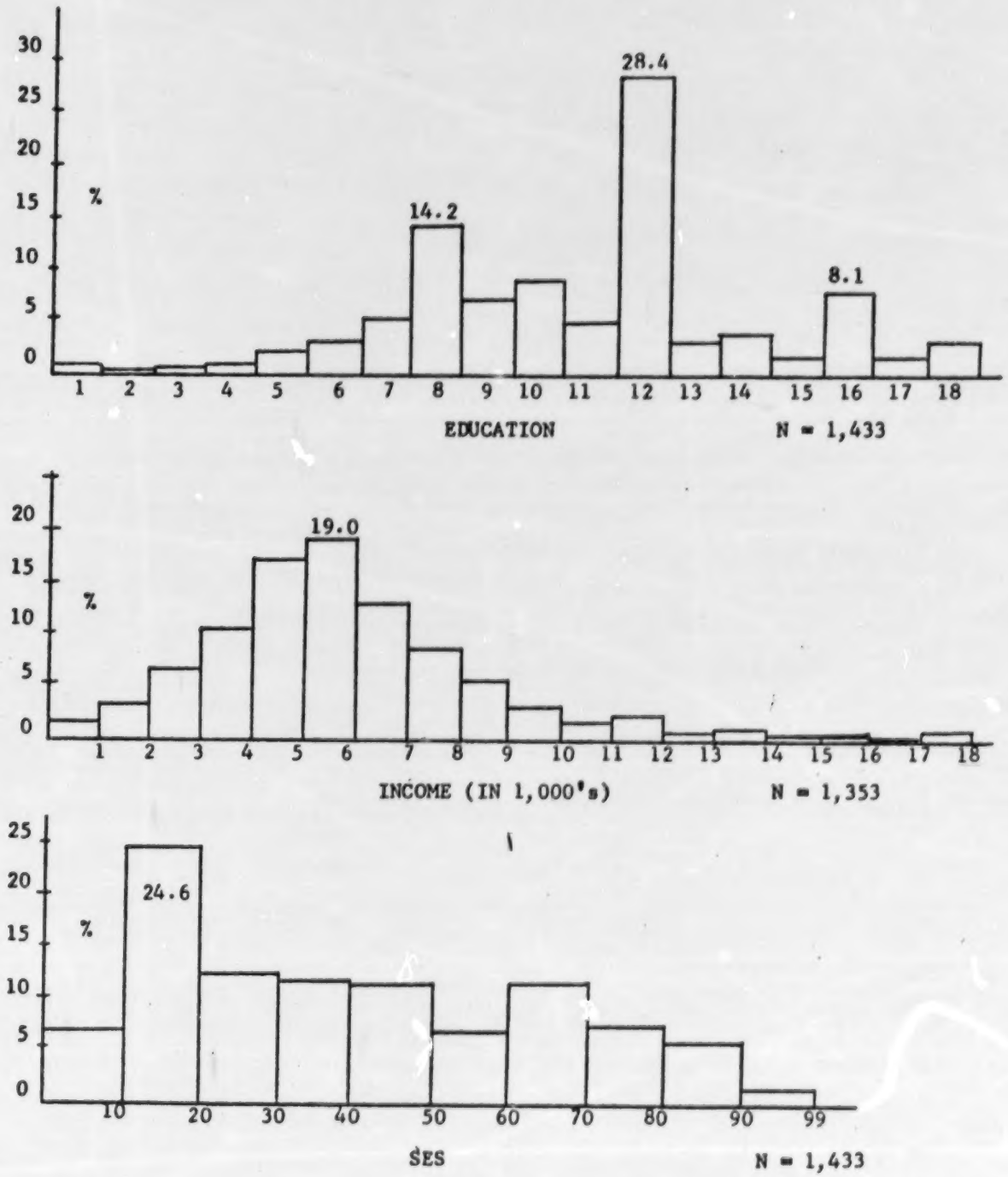

FIGURE 3. Bar Graphs of Distribution of Model Variables over Standard Population. (Source: CES Contingency Tables).

TABLE 1. Correlation and Regression Parameters for the Standard Population: Zero Order Correlations, Means, Standard Deviations; Standardized Regression (Path) Coeffients, Coefficient of Determination, Prediction Equation.

	SES	INC	MEAN	ST'D DEV
EDUC	.5250	.2967	10.82	3.42
SES		.3900	39.91	23.88
INC			6464.97	4870.82

N = 1,352

$P_{ses \leftarrow ed} = .52$; $P_{ses \leftarrow res} = .85$

$P_{inc \leftarrow ed} = .13$; $P_{inc \leftarrow ses} = .32$; $P_{inc \leftarrow res} = .89$

$R = .405$; $R^2 = .164$

Prediction equation: INC = 1940.52 + 180.99(EDUC) + 65.96(SES).

Range of prediction: $1,940.52 - $11,794.34

Source: CES.

TABLE 2. Socioeconomic Status x Income: CES Standard Population (Truncated).

Income (Dollars)	Socioeconomic Status*										
	0 - 9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	Total
0 - 1,000	4	9	4	2	0	1	4	1	0	0	25 .018
1,001 - 2,000	10	24	1	5	2	0	0	0	0	0	42 .031
2,001 - 3,000	19	35	10	11	4	2	4	1	2	1	89 .066
3,001 - 4,000	18	55	24	15	19	4	6	4	1	0	146 .108
4,001 - 5,000	24	76	44	18	28	17	11	12	8	0	238 .176
5,001 - 6,000	15	60	40	31	40	22	27	12	10	0	257 .190
6,001 - 7,000	8	33	28	27	22	21	21	10	7	1	178 .137
7,001 - 8,000	1	19	10	17	15	17	16	14	7	0	116 .086
8,001 - 9,000	0	12	4	10	9	11	14	9	7	2	78 .058
9,001 - 10,000	1	3	1	4	4	2	9	9	9	0	42 .031
10,001 - 11,000	0	3	2	3	1	3	5	4	6	0	27 .020
11,001 - 12,000	0	1	0	3	3	1	10	2	9	2	31 .023
12,001 - 13,000	0	0	1	1	1	0	2	5	4	0	14 .010
13,001 - 14,000	0	1	0	1	0	0	5	6	1	0	14 .010
14,001 - 15,000	0	1	0	0	0	0	2	1	4	0	8 .006
15,001 - 16,000	0	0	0	0	0	2	2	0	1	1	6 .004
16,001 - 17,000	0	1	0	0	0	0	1	0	0	0	2 .001
Total	100	333	169	148	148	103	139	90	76	7	1313

*Duncan socioeconomic index for detailed occupations.

TABLE 3. Education x Socioeconomic Status: CES Standard Population.

Socioeconomic Status*	Educational Attainment (Years)																		Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
0 - 9	3	1	3	2	6	7	14	18	13	8	6	20	2	0	0	0	0	0	103
																			.072
10 - 19	6	0	8	6	13	19	25	67	32	39	19	89	6	8	4	7	2	2	352
																			.246
20 - 29	1	1	0	6	3	7	20	34	18	17	11	45	7	1	0	0	0	0	175
																			.122
30 - 39	0	0	0	1	2	5	3	34	10	23	10	54	4	6	3	7	0	0	162
																			.113
40 - 49	0	0	0	1	1	5	5	27	14	14	8	51	15	6	4	6	2	0	159
																			.111
50 - 59	2	0	0	1	1	2	5	12	3	9	3	44	2	9	2	10	1	5	111
																			.077
60 - 69	1	0	0	0	1	1	3	10	8	11	6	53	8	13	5	28	3	8	159
																			.111
70 - 79	2	0	0	0	0	0	1	0	1	6	4	32	7	8	5	26	10	7	109
																			.076
80 - 89	3	0	0	0	0	0	1	2	1	3	3	15	2	6	6	30	6	8	86
																			.060
90 - 99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	12	17
																			.012
Total	18	2	11	17	27	46	77	204	100	130	70	407	53	57	30	116	26	42	1433
	.013	.001	.008	.012	.019	.032	.054	.142	.070	.091	.049	.284	.037	.040	.021	.081	.018	.029	

*Duncan socioeconomic index for detailed occupations.

TABLE 4. Education x Income: CES Standard Population (Truncated).

Income (Dollars)	Educational Attainment (Years)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
0 - 1,000	1	0	1	2	1	2	2	3	2	3	1	3	1	0	0	3	0	0	25 .018
1,001 - 2,000	0	1	3	1	2	6	4	6	6	3	1	4	2	1	1	0	0	1	42 .031
2,001 - 3,000	1	0	1	3	5	4	10	23	11	6	1	17	0	1	0	3	0	3	89 .066
3,001 - 4,000	3	0	2	2	5	7	8	25	16	17	9	35	4	3	1	5	2	2	146 .108
4,001 - 5,000	2	1	1	3	7	8	16	42	11	24	16	73	9	7	3	11	4	0	238 .176
5,001 - 6,000	2	0	2	2	3	12	14	37	25	22	16	81	10	10	1	16	3	1	257 .190
6,001 - 7,000	0	0	1	1	3	3	9	27	11	21	12	56	10	3	5	11	3	2	178 .132
7,001 - 8,000	1	0	0	1	1	0	7	12	3	12	6	45	4	7	4	8	3	2	116 .086
8,001 - 9,000	1	0	0	1	0	1	1	8	1	10	1	20	6	4	3	13	2	7	79 .058
9,001 - 10,000	0	0	0	0	0	0	0	3	3	4	2	12	1	3	1	5	3	5	42 .031
10,001 - 11,000	0	0	0	0	0	0	0	5	1	2	0	10	2	1	0	4	1	1	27 .020
11,001 - 12,000	0	0	0	0	0	1	1	1	1	0	0	11	1	0	3	7	1	4	31 .023
12,001 - 13,000	0	0	0	0	0	0	1	0	0	0	1	1	0	1	2	6	1	1	14 .010
13,001 - 14,000	1	0	0	0	0	0	0	2	2	0	0	4	0	3	1	0	0	1	14 .010
14,001 - 15,000	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	4	0	1	8 .006
15,001 - 16,000	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	1	1	6 .004
16,001 - 17,000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2 .001
Total	12	2	11	16	27	45	74	195	94	124	66	372	51	47	25	96	24	33	1314

TABLE 5. Low Earners - Income x Education for Three Standard Population Subsets: Full Time Workers, Workers Between 30 and 54 Years of Age and Full Time Workers Between 30 and 54 Years of Age. (Source: CES).

Income (Dollars)	Educational Attainment (Years)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Full Time Workers																			
0 - 1,000	1	0	1	2	0	1	1	0	1	1	0	2	0	0	0	1	0	0	11 .010
1,001 - 2,000	0	0	0	1	1	3	2	3	3	2	1	2	1	0	0	0	0	0	19 .017
2,001 - 3,000	1	0	0	2	4	2	5	13	5	5	1	15	0	1	0	2	0	1	57 .051
3,001 - 4,000	3	0	1	1	5	5	5	18	11	11	7	26	4	2	0	3	2	0	104 .094
Total	5	0	2	6	10	11	13	34	20	19	9	45	5	3	0	6	2	1	191
Workers, Age 30-54																			
0 - 1,000	1	0	1	1	1	2	1	3	2	3	1	1	0	0	0	3	0	0	20 .021
1,001 - 2,000	0	0	1	0	2	5	4	2	2	1	1	3	1	0	0	0	0	0	22 .023
2,001 - 3,000	0	0	1	3	3	3	6	12	5	2	1	7	0	0	0	1	0	1	45 .047
3,001 - 4,000	3	0	1	1	3	5	5	14	14	13	7	18	2	1	1	2	1	0	91 .095
Total	4	0	4	5	9	15	16	31	23	19	10	29	3	1	1	6	1	1	178
Full Time Workers Age 30-54																			
0 - 1,000	1	0	1	1	0	1	1	0	1	1	0	1	0	0	0	1	0	0	9 .011
1,001 - 2,000	0	0	0	0	1	2	2	2	0	0	1	1	1	0	0	0	0	0	10 .012
2,001 - 3,000	0	0	0	2	2	1	2	6	2	2	1	5	0	0	0	1	0	1	25 .031
3,001 - 4,000	3	0	1	1	3	3	3	10	10	9	5	14	2	1	0	2	1	0	68 .085
Total	4	0	2	4	6	7	8	18	13	12	7	21	3	1	0	4	1	1	112